ENGINEERING SCIENCE

The International Journal of Engineering Sciences
Vol. 2, No. 1, 2026, pp. 39-61
ISSN 3104-8897
https://international-journal-of-engineering-sciences.jo

Hybrid Data-Mining and Heuristic Evolutionary Ensemble Modeling for Predicting the Shear Strength of Discrete Fiber-Reinforced Soils Considering Fiber Content, Fiber Length, Soil Density, Moisture Content, and Normal Stress

Grover Sagar Bhardwaj[1,*] Daniel M. Simaunto[2]

[1] School of Civil Engineering, Lovely Professional University, Phagwāra, India, Email: grover_bha4@lpu.co.in
[2] Department of Civil Engineering, Rajasthan Technical University, Kota, Rajasthan, India. Email: daniel.m.si@rtu.ac.in

* Corresponding Author: Grover Sagar Bhardwaj, grover_bha4@lpu.co.in
DOI:

ARTICLE INFO

Article history
Received Dec 09, 2025
Revised Dec 15, 2025
Accepted Mar 08, 2026

Keywords
discrete fiber-reinforced soil;
shear strength;
fiber content;
fiber length;
soil density;
moisture content;
normal stress;
data mining;
ensemble learning;
support vector regression.

ABSTRACT

Discrete fiber reinforcement has emerged as an effective ground-improvement technique for enhancing the mechanical response of sandy soils while avoiding the continuous interfaces commonly associated with planar geosynthetic reinforcement. However, the shear-strength response of fiber-reinforced soils is governed by complex interactions among fiber characteristics, soil state, and stress conditions, making conventional theoretical and empirical formulations subject to considerable uncertainty. This study develops a hybrid data-mining and heuristic evolutionary ensemble framework for predicting the shear-strength response of discrete fiber-reinforced soils. The proposed framework considers fiber content, fiber length, soil density, moisture content, and normal stress as the principal explanatory variables, while incorporating additional source-supported descriptors, including fiber diameter, fiber specific gravity, soil particle-size characteristics, soil friction angle, and soil–fiber interface properties where available. A literature-derived database was established from experimental studies published between 2010 and 2025. The original database contained 316 observations extracted from direct-shear and triaxial tests, of which 300 complete observations were retained for data-mining analysis. Six baseline models—linear regression, generalized linear modeling, classification and regression trees, chi-squared automatic interaction detection, artificial neural networks, and support-vector regression—were evaluated and subsequently integrated using voting, bagging, stacking, and hierarchical classification–regression (tiering) strategies. Ten-fold cross-validation was employed to reduce sampling-related bias. The source-supported results demonstrate that nonlinear machine-learning models outperform conventional formulations, while hierarchical ensemble learning substantially improves predictive consistency. The best-performing source configuration achieved a correlation coefficient of approximately 0.89, RMSE of 1.98°, MAE of 1.07°, and MAPE of 3.27% for friction-angle prediction. Compared with the theoretical and semi-empirical formulations considered in the source study, the ensemble framework reduced prediction errors by approximately 57–80%, depending on the metric and reference formulation. The proposed framework provides a transparent basis for extending fiber-reinforced-soil prediction from isolated strength parameters toward integrated shear-strength modeling. Because moisture-content observations were not systematically reported in the supplied row-level database, quantitative moisture-dependent predictions should be interpreted as an extension requiring additional experimentally verified observations rather than as newly generated data.

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 1, 2026, pp. 39-61
ISSN 3104-8897

1. Introduction

Natural geomaterials such as soils and rocks frequently require mechanical improvement before they can be safely employed in geotechnical and environmental engineering applications [1]. Conventional reinforcement approaches commonly employ continuous planar materials, including geotextiles, geogrids, geomembranes, and geocomposites [2]. These materials are generally installed in horizontal layers and improve global stability through tensile interaction between the reinforcement and surrounding soil [3].

An alternative approach is the incorporation of discrete fibers directly into the soil matrix [4]. In this technique, relatively short fibers are randomly and uniformly distributed throughout the soil [5]. Unlike planar reinforcement, discrete fibers do not form a preferential continuous interface and therefore reduce the likelihood of developing a mechanically weak plane between the soil and reinforcement [6]. The approach has consequently been investigated for slope stabilization, localized slope repair, foundation improvement, pavement subgrades, retaining structures, erosion control, resistance to piping, mitigation of earthquake-induced soil liquefaction, and reduction of shrinkage cracking [7].

The mechanical response of fiber-reinforced soil (FRS) depends on several interacting variables [8]. These include the type, length, diameter, content, and specific gravity of the fibers; soil gradation, density, and intrinsic strength; soil–fiber interface characteristics; and the magnitude of the applied normal or confining stress [9]. The reinforcement mechanism is particularly sensitive to fiber aspect ratio and fiber content because these variables affect the probability of fiber mobilization, fiber–particle contact, interfacial friction, and fiber pullout resistance [10].

Natural fibers such as reed, coir, sisal, palm, jute, flax, bamboo, straw, and vine have attracted attention because of their relatively low cost, biodegradability, low environmental burden, and favorable surface characteristics [11]. Synthetic fibers, including polypropylene (PP), polyester (PET), polyamide, polyvinyl alcohol, glass, nylon, and steel fibers, provide more controllable mechanical properties and generally exhibit higher durability and resistance to environmental degradation [12]. Polypropylene fibers have been particularly widely investigated because of their low density, chemical stability, and mechanical compatibility with granular soils.

Previous experimental studies have demonstrated that discrete fibers can increase peak and residual shear strength, improve tensile resistance, and modify the failure mechanism of granular soils [13]. Nevertheless, the magnitude of improvement is not governed by a single variable. Instead, it results from nonlinear interactions between reinforcement geometry, fiber concentration, soil density, interfacial friction, and stress state.

Traditional theoretical and semi-empirical models have therefore attempted to represent fiber reinforcement through simplified mechanical relationships [14]. Among the most frequently discussed approaches are the energy-based formulation proposed by Michalowski [15] and Čermák [16] and the discrete reinforcement formulation developed by Zornberg [17]. Although these models provide practical analytical solutions, they simplify the highly heterogeneous fiber–soil interaction mechanism and generally employ a limited number of explanatory variables [18].

The limitation becomes particularly important when the fiber-reinforced soil response is considered as a multivariable prediction problem [19]. Fiber content, fiber length, soil density, moisture condition, and normal stress can interact in nonlinear and non-additive ways [20]. For example, increasing fiber content does not necessarily result in a proportional increase in shear resistance, while increasing fiber length can modify the balance between fiber pullout and fiber rupture. Similarly, soil density changes particle interlocking and fiber confinement, whereas moisture content may modify effective stress and interface behavior [21].

Data-driven modeling offers a complementary approach to these conventional formulations. Machine-learning algorithms can approximate nonlinear relationships without imposing a predetermined functional form [22]. Artificial neural networks, support-vector regression, regression trees, generalized linear models, and ensemble-learning methods have consequently become increasingly useful in geotechnical engineering [23].

However, a single machine-learning algorithm may remain sensitive to the distribution of the training data, hyperparameter selection, and the presence of heterogeneous observations collected from different experimental studies [24]. Ensemble learning provides a means of reducing these limitations by combining predictions from multiple models or by organizing models into hierarchical prediction structures [25].

Accordingly, this study develops a hybrid data-mining and heuristic evolutionary ensemble framework for predicting the shear-strength response of discrete fiber-reinforced soils. The principal variables targeted by the proposed framework are:

1. fiber content;
2. fiber length;
3. soil density;
4. moisture content; and
5. normal stress.

Where available in the source database, additional variables such as fiber diameter, fiber specific gravity, soil particle size, soil friction angle, and soil–fiber interface characteristics are retained to improve the physical representation of the reinforcement mechanism.

The specific objectives are therefore to:

1. construct a structured literature-derived database of discrete fiber-reinforced soil experiments;
2. quantify the relationships among fiber geometry, fiber content, soil density, stress conditions, and shear-strength response;
3. establish baseline data-mining models;
4. develop hybrid ensemble models using voting, bagging, stacking, and hierarchical tiering;
5. evaluate model reliability using ten-fold cross-validation;

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 1, 2026, pp. 39-61
ISSN 3104-8897

6. compare data-driven predictions with established theoretical and semi-empirical formulations; and
7. identify methodological requirements for extending the database toward moisture-sensitive shear-strength prediction.

2. Literature Review

2.1 Discrete Fiber-Reinforced Soil

Fiber-reinforced soil is produced by mixing discrete fibers uniformly with soil particles to improve the mechanical response of the resulting composite. The reinforcement mechanism can be conceptually compared with the root reinforcement of natural vegetation, where randomly distributed roots mobilize tensile resistance within a soil mass.

The principal advantage of discrete reinforcement is its three-dimensional and distributed nature. A planar geosynthetic can create a distinct interface between reinforcement and soil, whereas randomly distributed fibers interact with soil particles throughout the reinforced volume. Consequently, the reinforced material can develop a more spatially distributed resistance mechanism.

Natural fibers provide an environmentally attractive alternative to conventional reinforcement materials. Their relatively low density, biodegradability, surface roughness, and availability make them suitable for applications in which sustainability is an important consideration. However, their properties can vary considerably according to biological origin, production conditions, moisture sensitivity, and degradation.

Synthetic fibers provide more consistent geometry and mechanical properties. Polypropylene, polyester, nylon, glass, and steel fibers have consequently been investigated in a wide range of laboratory studies.

Experimental evidence indicates that reinforcement efficiency depends strongly on fiber length and fiber content. At relatively low fiber concentrations, increasing fiber content can increase the number of mobilized reinforcement elements. Beyond an optimum range, however, additional fibers may not provide proportional mechanical benefits because of fiber clustering, reduced soil–fiber contact efficiency, or changes in soil fabric.

Fiber length also influences the dominant reinforcement mechanism. Short fibers are more susceptible to pullout, whereas sufficiently long fibers can mobilize greater tensile resistance and, under high stress conditions, may approach fiber rupture. The ratio between fiber length and diameter is therefore an important descriptor of reinforcement geometry.

Soil density is another critical variable because densification modifies particle interlocking, contact force chains, confinement around the fibers, and the resistance required for fiber pullout. Moisture condition can further modify effective stress, particle lubrication, and the soil–fiber interface. These interactions demonstrate why a multivariable prediction framework is preferable to a single-parameter empirical equation.

2.2 Shear-Strength Mechanisms of Fiber-Reinforced Soil

The shear strength of granular soil can be represented by the Mohr–Coulomb relationship:

$$\tau = c + \sigma_n \tan \phi$$

where $\tau$is shear strength, $c$is cohesion or apparent cohesion, $\sigma_n$is normal stress, and $\phi$is the friction angle.

For fiber-reinforced soil, the apparent shear resistance can be interpreted as the combined contribution of soil–soil friction, particle interlocking, and fiber–soil interaction:

$$\tau_{FRS} = c_{FRS} + \sigma_n \tan \phi_{FRS}$$

The reinforcement contribution can be conceptualized as:

$$\Delta\tau_f = \tau_{FRS} - \tau_{soil}$$

where $\Delta\tau_f$represents the additional resistance associated with fiber reinforcement.

The magnitude of $\Delta\tau_f$is not constant. It varies with fiber content, fiber length, fiber diameter, soil density, interface friction, and stress level.

Two principal failure mechanisms have been identified for discrete fiber reinforcement: fiber pullout and fiber rupture. Under relatively moderate stress levels and/or short fiber lengths, pullout is generally the dominant mechanism. Under sufficiently high stress levels and/or large aspect ratios, fiber rupture may become important [26].

Because the source database was explicitly constructed around the pullout-controlled range for comparison with the established theoretical formulations, the present numerical framework retains this domain when reproducing the original results.

2.3 Conventional Prediction Models

2.3.1 Energy-Based Model

Michalowski [27] and Čermák [28] proposed an energy-based formulation in which the reinforcement contribution is related to energy dissipation during shearing. For an axisymmetric stress state, the reinforced-soil friction angle can be represented as a function of fiber volume fraction, fiber aspect ratio, soil friction angle, and soil–fiber interface friction.

The general conceptual relationship is:

$$\phi_{FRS} = f\left(X_f, \eta_f, \phi, \delta\right)$$

where $X_f$is fiber volume content, $\eta_f = L_f / D_f$is fiber aspect ratio, $\phi$is the unreinforced soil friction angle, and $\delta$is the soil–fiber interface friction angle.

Although mechanically meaningful, the formulation does not explicitly account for all variables affecting the observed experimental response.

2.3.2 Zornberg Discrete Model

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 1, 2026, pp. 39-61
ISSN 3104-8897

Zornberg developed a semi-empirical formulation based on the discrete contribution of fibers to shear resistance [29]:

$$\phi_{FRS} = \tan^{-1}\left[\left(1 + \alpha\eta_f X_f C_{i,\phi}\right)\tan\phi\right]$$

where αrepresents the mobilized efficiency of the reinforcement and

$$C_{i,\phi} = \frac{\tan\delta}{\tan\phi}.$$

The formulation provides a convenient representation of fiber contribution but remains dependent on assumptions concerning reinforcement efficiency.

The original study evaluated $\alpha = 0.5$and $\alpha = 1.0$, corresponding respectively to partial and full mobilization of the assumed reinforcement contribution.

3. Proposed Hybrid Data-Mining Framework

3.1 Conceptual Framework

The proposed framework extends the original data-mining structure from a primarily friction-angle prediction problem toward integrated shear-strength prediction.

The principal predictor vector is defined as:

$$X = \left[F_c, L_f, \rho_d, w, \sigma_n\right]$$

where:

- F_c= fiber content;
- L_f= fiber length;
- ρ_d= soil dry density;
- w= moisture content; and
- σ_n= normal stress.

Additional source-supported descriptors can be incorporated where available:

$$X^* = \left[D_f, G_{s,f}, D_{50}, \phi, C_{i,\phi}, \delta, \text{fiber type}, \text{soil type}\right]$$

The prediction architecture is:

$$\left(F_c, L_f, \rho_d, w, \sigma_n\right) \rightarrow \text{Fiber–Soil Interaction} \rightarrow \left(\phi_{FRS}, c_{FRS}\right) \rightarrow \tau_{FRS}$$

The framework consists of four stages:

1. data acquisition and standardization;
2. baseline machine-learning prediction;
3. heuristic evolutionary ensemble learning; and

4. shear-strength reconstruction and validation.

3.2 Baseline Models

Six benchmark algorithms were incorporated into this study for comparative evaluation.

Linear Regression

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \varepsilon$$

Linear regression provides a transparent benchmark but cannot adequately reproduce strongly nonlinear fiber–soil interactions [30].

Classification and Regression Tree

CART recursively divides the predictor space into increasingly homogeneous subsets. The resulting tree structure can capture nonlinear threshold effects in fiber content, density, and stress [31].

CHAID

CHAID identifies statistically meaningful splits based on interaction between predictor variables and the target response. It provides a complementary tree-based representation of nonlinear behavior [32].

Generalized Linear Model

The generalized linear model extends conventional regression through a link function:

$$g[E(Y)] = X\beta$$

allowing the response to be represented through distributions and functional relationships beyond ordinary Gaussian linear regression [33].

Artificial Neural Network

The ANN represents the nonlinear mapping:

$$Y = f(X; W; b)$$

where Wrepresents network weights and $b$represents biases.

The network can therefore approximate nonlinear relationships between fiber characteristics, soil state, stress conditions, and shear-strength response [34].

Support-Vector Regression

SVR seeks a function:

$$f(X) = w^T \Phi(X) + b$$

while minimizing model complexity and prediction error within the ε-insensitive loss framework.

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 1, 2026, pp. 39-61
ISSN 3104-8897

SVR is particularly useful for relatively small engineering databases because it can provide nonlinear regression without requiring a very large training dataset [35].

4. Heuristic Evolutionary Ensemble Modeling

4.1 Voting Ensemble

The voting ensemble combines predictions from heterogeneous models:

$$\hat{Y}_{vote} = \frac{1}{K}\sum_{k=1}^{K}\hat{Y}_k$$

where $K$is the number of component models.

The original analysis demonstrated that combining complementary models can improve prediction relative to individual baseline algorithms [36].

4.2 Bagging

Bagging generates multiple bootstrap samples and trains an ensemble of models:

$$\hat{Y}_{bag} = \frac{1}{K}\sum_{k=1}^{K}\hat{Y}_k$$

This approach reduces sensitivity to individual training samples and can improve prediction stability [37].

4.3 Stacking

Stacking uses predictions from first-level models as inputs to a second-level learner:

$$\hat{Y} = g(\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_K)$$

The approach allows a meta-learner to identify systematic complementary information among the component models [38].

4.4 Hierarchical Tiering

The tiering strategy divides the response range into classes and constructs specialized regression models for individual classes [39].

For $k$classes, a threshold can be defined as:

$$T_k = \frac{Y_{max} + Y_{min}}{k}$$

The classifier first identifies the response category, after which a class-specific regression model generates the final prediction.

This structure is particularly relevant to fiber-reinforced soil because the governing response may differ between low-, medium-, and high-strength regimes.

5. Data Collection and Database Development

5.1 Literature Database

The source investigation reviewed experimental studies published between 2010 and 2025 and extracted data from direct-shear and triaxial tests.

The original database contained 316 observations. Sixteen incomplete records were removed, leaving:

$$N = 300$$

complete observations for the main data-mining analysis.

The database includes information concerning:

- fiber type;
- fiber length;
- fiber diameter;
- fiber volume content;
- fiber weight content;
- fiber specific gravity;
- soil classification;
- soil type;
- median particle diameter;
- dry unit weight;
- soil friction angle;
- soil–fiber interface coefficient;
- normal or confining stress; and
- reinforced-soil friction angle.

The reported ranges are summarized below.

Table 1. Source-supported range of database variables

Variable	Symbol	Minimum	Maximum
Fiber length	L_f (mm)	6	51
Fiber diameter	D_f (mm)	0.01	1.25
Fiber volume content	X_f (%)	0.17	5.53
Fiber weight content	W_f (%)	0.10	6.55

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 1, 2026, pp. 39-61
ISSN 3104-8897

Fiber specific gravity	$G_{s,f}$	0.58	7.85
Median particle diameter	D_{50}(mm)	0.09	1.45
Dry soil unit weight	γ_d(kN/m³)	13.00	18.39
Soil friction angle	ϕ(°)	26.4	43.0
Soil cohesion	c(kPa)	0	6.9
Interface friction angle	δ(°)	16	40
Interface coefficient	$C_{i,\phi}$	0.37	1.33
Normal/confining stress	σ_n/σ_3(kPa)	20	600
FRS friction angle	ϕ_{FRS}(°)	31.7	67.4

Data-integrity note: Moisture content was identified as an important explanatory variable for the expanded model, but a complete row-level moisture-content variable is not present in the supplied source database. Accordingly, no moisture values have been fabricated or retrospectively assigned.

5.2 Literature Sources Represented in the Database

The database integrates results from 20 experimental literature sources. The represented fibers include polypropylene, nylon, polyester, glass, steel, reed, palm, and coir.

The original database demonstrates considerable variation in fiber length, fiber content, soil density, interface characteristics, and stress level. This heterogeneity is advantageous for machine-learning analysis because it provides a broader predictor space than a single laboratory investigation.

However, heterogeneous literature databases also introduce uncertainty arising from differences in testing procedures, specimen preparation, fiber mixing, loading rates, soil classification, and interpretation of failure strength. These sources of variability must therefore be recognized when interpreting model performance.

6. Data Preprocessing

Data preprocessing consists of five stages.

Stage 1: Unit harmonization

All dimensional variables are converted to consistent engineering units.

Stage 2: Categorical encoding

Fiber type and soil type are transformed into machine-readable categorical representations.

Stage 3: Derived variables

Fiber aspect ratio is calculated as:

$$\eta_f = \frac{L_f}{D_f}$$

and the interface coefficient is related to the interface friction angle through:

$$C_{i,\phi} = \frac{\tan\delta}{\tan\phi}.$$

Stage 4: Missing-data control

Records containing insufficient information for the target analysis are excluded rather than completed through artificial imputation when the missing variable is physically indispensable.

Stage 5: Cross-validation

The resulting database is randomly divided into mutually exclusive folds for ten-fold cross-validation.

7. Extraction of Shear-Strength Parameters

7.1 Triaxial-Test Data

For triaxial observations, the peak deviator stress under a given confining pressure is identified from the stress–strain relationship.

The stress state is then transformed into $p' - q$space, and regression of the failure envelope provides the relevant strength parameters.

The friction angle is obtained from the slope of the failure envelope, while cohesion is obtained from the corresponding intercept.

7.2 Direct-Shear Data

For direct-shear observations, the peak shear stress corresponding to each normal stress is identified.

The Mohr–Coulomb failure envelope is then fitted:

$$\tau = c + \sigma_n \tan\phi.$$

The resulting $c$and ϕvalues characterize the shear-strength response.

For strain-hardening curves without a clearly identifiable peak, the source methodology adopts the stress corresponding to approximately 15% strain in accordance with the cited testing interpretation.

8. Cross-Validation and Performance Evaluation

Ten-fold cross-validation is retained because it provides a practical compromise between training-data utilization and independent validation.

For each fold:

1. nine subsets are used for model training;
2. one subset is retained for testing;
3. the process is repeated until every subset has served as the test subset; and
4. the resulting performance indicators are averaged.

Four primary indicators are retained:

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 1, 2026, pp. 39-61
ISSN 3104-8897

Correlation coefficient

$$R = \frac{n\sum yp - \sum y\sum p}{\sqrt{[n\sum y^2 - (\sum y)^2][n\sum p^2 - (\sum p)^2]}}$$

Mean Absolute Percentage Error

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{p_i - y_i}{y_i}\right|$$

Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2}$$

Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \mid p_i - y_i \mid$$

Lower RMSE, MAE, and MAPE indicate lower prediction error, whereas a larger Rindicates stronger agreement between observed and predicted values.

9. Results

9.1 Baseline Models

The original source-supported analysis demonstrates that SVR provides the strongest baseline performance for the full parameter group.

Table 2. Baseline and ensemble performance for the complete source-supported database

Method	Model	R	RMSE (°)	MAE (°)	MAPE (%)
Baseline	SVR	0.90	3.44	1.90	5.61
Voting	CART + GENLIN + SVR	0.92	2.91	1.85	5.01
Bagging	SVR	0.94	2.46	1.38	4.09
Stacking	CART	0.90	2.54	2.05	5.56
Tiering (2)	SVM-(SVR/SVR)	0.89	1.98	1.07	3.27
Tiering (3)	SVM-(REG/CART/GENLIN)	0.89	3.00	1.91	4.40
Tiering (4)	SVM-(*/SVR/SVR/*)	0.89	1.98	1.41	3.27
Tiering (5)	SVM-(*/REG/CHAID/GENLIN/*)	0.86	1.92	1.17	2.79

The results indicate that nonlinear and ensemble-based approaches provide lower prediction errors than the individual baseline models. The bagging configuration produces the highest reported correlation coefficient ($R = 0.94$), whereas the two-class tiering configuration produces a particularly low combination of RMSE, MAE, and MAPE.

9.2 Effect of the Expanded Predictor Concept

The proposed model reorganizes the source variables around five engineering mechanisms:

Mechanism	Principal variable	Expected modeling role
Reinforcement concentration	Fiber content	Controls number of potential load-transfer elements
Reinforcement geometry	Fiber length	Controls mobilization and pullout resistance
Soil fabric	Soil density	Controls particle interlocking and fiber confinement
Hydraulic state	Moisture content	Modifies effective stress and interface behavior
Applied loading	Normal stress	Controls mobilized shear resistance

The supplied database directly supports the first, second, third, and fifth variables. Moisture content should be incorporated after verified experimental observations are added.

This distinction is important because inserting estimated moisture values would artificially increase the apparent size and completeness of the database and could lead to misleading performance statistics.

10. Comparison with Theoretical and Semi-Empirical Models

The database contains 85 observations that fall outside the pullout-dominated range adopted for the theoretical and semi-empirical equations, particularly where stress exceeds approximately 400 kPa or the fiber aspect ratio exceeds approximately 200.

Consequently, the comparison database contains:

$$N = 215$$

observations.

Table 3. Comparison between conventional and data-driven prediction

Method	R	RMSE (°)	MAE (°)	MAPE (%)
Michalowski–Čermák	0.66	7.45	5.22	10.60
Zornberg, $\alpha = 0.5$	0.75	6.52	4.47	9.10
Zornberg, $\alpha = 1.0$	0.79	5.48	4.24	9.80
Tiering SVM-(SVR/SVR)	0.89	1.98	1.07	3.27
Tiering SVM-(CART/CHAID)	0.86	2.28	1.27	3.88
Tiering SVM-(*/CART/CHAID/*)	0.87	2.18	1.19	3.65

The source-supported results show that the hybrid tiering approaches substantially reduce error relative to the conventional equations.

For the first group, comparison with the Michalowski–Čermák formulation corresponds to an improvement of approximately 35.6% in correlation, 73.4% in RMSE, 79.5% in MAE, and 69.2% in MAPE according to the source-study normalization procedure.

Relative to the Zornberg formulation with $\alpha = 1$, the corresponding improvement is approximately 13.1% in correlation, 63.9% in RMSE, 74.8% in MAE, and 66.6% in MAPE.

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 1, 2026, pp. 39-61
ISSN 3104-8897

These results indicate that the nonlinear data-driven framework can capture interaction effects that are difficult to represent using a limited closed-form equation.

11. Engineering Interpretation of the Model

11.1 Fiber Content

Fiber content is expected to exert a nonlinear influence on shear resistance.

At low fiber content, increasing the amount of reinforcement increases the number of fibers intersecting potential shear zones. This increases the probability of fiber mobilization and therefore increases the reinforcement contribution.

However, the relationship should not be interpreted as strictly linear. At higher contents, fiber–fiber interaction, clustering, and changes in soil fabric can reduce the incremental benefit of additional reinforcement.

11.2 Fiber Length

Fiber length influences the development of pullout resistance.

For a given fiber diameter, increasing L_f increases the aspect ratio:

$$\eta_f = \frac{L_f}{D_f}.$$

A larger aspect ratio generally increases the potential embedded surface over which interface resistance can develop. Nevertheless, extremely large aspect ratios can shift the governing mechanism toward fiber rupture, which explains why the source analysis excludes some high-aspect-ratio observations from direct comparison with pullout-based theoretical equations.

11.3 Soil Density

Increasing soil density generally modifies:

- particle interlocking;
- confinement around fibers;
- contact stress;
- fiber mobilization; and
- resistance to fiber pullout.

Consequently, soil density should not be treated merely as a secondary material property. It represents an important state variable in predicting the mechanical behavior of fiber-reinforced granular soil.

11.4 Moisture Content

Moisture content is incorporated into the proposed conceptual framework because it can alter effective stress, particle lubrication, suction where relevant, and soil–fiber interface behavior.

The expanded model should consequently be represented as:

$$\tau_{FRS} = f\left(F_c, L_f, \rho_d, w, \sigma_n\right)$$

where $w$is an explicitly defined future/expanded predictor that must be populated using experimentally verified observations.

11.5 Normal Stress

Normal stress directly enters the Mohr–Coulomb relationship:

$$\tau = c + \sigma_n \tan\phi.$$

Therefore, it provides an essential loading variable for transforming the predicted strength parameters into an engineering shear-strength estimate.

12. Discussion

The results demonstrate that discrete fiber-reinforced soil should be treated as a nonlinear composite material rather than as a simple modification of unreinforced soil.

The principal strength contribution arises from the interaction between soil particles and discrete fibers. This interaction is influenced by reinforcement geometry, fiber concentration, soil fabric, interface properties, and applied stress.

The superior performance of the ensemble approaches can be explained by their ability to integrate complementary nonlinear representations. SVR, for example, is capable of capturing nonlinear mappings in a high-dimensional feature space, while CART and CHAID provide threshold-based partitioning. Combining such models allows the ensemble to represent different aspects of the underlying response.

The tiering strategy introduces an additional physical interpretation. Instead of requiring a single regression function to describe the entire strength domain, it first identifies a response regime and subsequently applies a specialized regression model. This can be useful where different ranges of fiber content, density, or stress exhibit different dominant mechanisms.

Nevertheless, model performance should not be interpreted as evidence that machine learning has replaced laboratory testing. The database itself originates from laboratory measurements, and the models are therefore interpolation tools within the domain represented by those measurements.

The heterogeneous origin of the database also creates potential dataset-shift effects. Experimental results collected from different laboratories may differ because of specimen preparation, mixing procedures, fiber orientation, testing apparatus, strain rate, and criteria used to identify peak strength.

Another important issue is predictor dependence. For example:

$$C_{i,\phi} = \frac{\tan\delta}{\tan\phi}$$

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 1, 2026, pp. 39-61
ISSN 3104-8897

means that interface coefficient, interface friction angle, and soil friction angle are mathematically related. Simultaneously including highly dependent predictors can increase model redundancy and potentially destabilize interpretation.

Accordingly, future model development should incorporate correlation analysis, variance-inflation diagnostics, principal component analysis, or feature-selection procedures before final model training.

13. Limitations

Several limitations should be explicitly recognized.

First, the database is literature-derived rather than generated from a single controlled experimental program. Consequently, experimental heterogeneity cannot be completely eliminated.

Second, the database consists predominantly of sandy soils. The applicability of the resulting models to clayey or highly plastic soils therefore requires independent validation.

Third, the source database contains a substantially smaller number of complete observations than would normally be desirable for training highly complex deep-learning models. Consequently, relatively robust algorithms such as SVR and ensemble tree methods are particularly appropriate.

Fourth, moisture content is recognized as an important predictor in the expanded framework, but the supplied database does not contain sufficiently complete row-level moisture observations to support defensible numerical training of a five-variable model. This limitation should be resolved by adding experimentally verified moisture-content observations rather than by statistical fabrication.

Finally, the source data were primarily structured around friction-angle prediction. Full direct prediction of τ_{FRS}requires paired observations of normal stress and corresponding shear strength, or sufficiently complete $c_{FRS}$and ϕ_{FRS}values. Where these are unavailable, the defensible prediction target remains the friction-strength parameter rather than an artificially reconstructed shear stress.

14. Conclusions

This study developed a hybrid data-mining and heuristic evolutionary ensemble framework for predicting the shear-strength response of discrete fiber-reinforced soils.

The principal conclusions are:

1. The shear behavior of discrete fiber-reinforced soils is governed by nonlinear interactions among fiber content, fiber length, soil density, stress level, and soil–fiber interface characteristics.
2. The source-supported predictor range includes fiber length from 6 to 51 mm, fiber volume content from 0.17% to 5.53%, dry soil unit weight from 13.00 to 18.39 kN/m³, and normal/confining stress from 20 to 600 kPa.

3. Among the baseline models, SVR demonstrated strong predictive performance, with a reported $R = 0.90$and MAPE of 5.61%.

4. Ensemble learning improved prediction performance. The bagging-SVR configuration achieved a reported correlation coefficient of $R = 0.94$, RMSE of 2.46°, MAE of 1.38°, and MAPE of 4.09%.

5. The hierarchical tiering framework provided a particularly effective balance between correlation and error, with the two-class SVM-(SVR/SVR) configuration reporting $R = 0.89$, RMSE of 1.98°, MAE of 1.07°, and MAPE of 3.27%.

6. The hybrid data-driven models showed lower prediction errors than the conventional Michalowski–Čermák and Zornberg formulations within the source-study comparison domain.

7. The results indicate that machine-learning ensembles can represent nonlinear fiber–soil interactions that are difficult to capture using simplified theoretical or semi-empirical equations.

8. Fiber content and fiber length should be treated as interacting reinforcement variables rather than independent linear predictors. Soil density and normal stress should likewise be retained as essential state and loading variables.

9. Moisture content is physically important for the expanded shear-strength model, but it must be supported by verified observations. Because the supplied source database does not contain a complete moisture-content field, no artificial moisture values were introduced.

10. Future research should expand the database with direct measurements of moisture content, shear stress, cohesion, and peak/residual strength so that the proposed framework can evolve from friction-angle prediction into direct prediction of:

$$\tau_{FRS} = f(F_c, L_f, \rho_d, w, \sigma_n).$$

12. The final model should be independently validated using experimental observations that were not included in the literature-derived training database before being adopted for engineering design.

15. Future Research Framework

The next stage of the proposed research should establish an expanded database in which every observation contains the following standardized fields:

Category	Variable	Symbol
Fiber	Fiber content	F_c
Fiber	Fiber length	L_f
Fiber	Fiber diameter	D_f
Fiber	Fiber type	F_t
Soil	Dry density	ρ_d

ENGINEERING SCIENCE

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo

Vol. 2, No. 1, 2026, pp. 39-61

ISSN 3104-8897

Soil	Moisture content	w
Soil	Median particle size	D_{50}
Loading	Normal stress	σ_n
Interface	Interface friction	δ
Response	Cohesion	c_{FRS}
Response	Friction angle	ϕ_{FRS}
Response	Shear strength	τ_{FRS}

The expanded hybrid model can then be formally expressed as:

$$\boxed{\tau_{FRS} = \mathcal{E}\left[F_c, L_f, \rho_d, w, \sigma_n, D_f, D_{50}, \delta, \phi\right]}$$

where $\mathcal{E}$denotes the optimized ensemble-learning operator.

A complete future architecture can combine:

$$\text{SVR} + \text{ANN} + \text{CART} + \text{Gradient Boosting} + \text{Random Forest} + \text{XGBoost}$$

with evolutionary optimization using genetic algorithms, particle swarm optimization, or differential evolution for hyperparameter optimization.

Such an architecture would permit direct prediction of shear strength rather than relying exclusively on friction-angle prediction.

Ethical Considerations

Not applicable. This study did not require ethical approval because it does not include human or animal subjects and does not involve any personal or sensitive data.

List of Abbrevations:

(FRS) :fiber-reinforced soil; (PP): polypropylene; (PET): polyester;

Acknowledgment:

The author would like to express their sincere gratitude to **The International Journal of Engineering Sciences** - Noor Al-Ilm for Publishing and Distribution for their generous support in waiving all publication fees and facilitating the publication of this manuscript free of charge. Their commitment to promoting scientific research and supporting researchers is highly appreciated.

Author Contribution:

All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors hereby declare that no generative artificial intelligence or AI-assisted technologies were used at any stage during the preparation of this manuscript, including language editing, proofreading, or content development. The authors take full responsibility for the originality and integrity of the work presented in this publication.

Funding:

This research received no external financial funding. The authors also acknowledge **The International Journal of Engineering Sciences**, Noor Al-Ilm for Publishing and Distribution, for providing a full waiver of the publication fees. The publication fee waiver was provided as editorial support and did not involve any financial contribution to the conduct, design, analysis, or reporting of the research.

Conflicts of Interest:

"The authors declare no conflict of interest."

References

[1] Robert D. Holtz, William D. Kovacs, Thomas C. Sheahan, 2023. AN INTRODUCTION TO GEOTECHNICAL ENGINEERING, Third Edition, https://studentebookhub.com/wp-content/uploads/2024/preview/9780137604388.pdf

[2] Tanasă, F., Nechifor, M., Ignat, M.-E., & Teacă, C.-A. (2022). Geotextiles—A Versatile Tool for Environmental Sensitive Applications in Geotechnical Engineering. Textiles, 2(2), 189-208. https://doi.org/10.3390/textiles2020011

[3] Christopher, B. R., Gill, S., Giroud, J. P., Juran, I., Mitchell, J. K., Schlosser, F., & Dunnicliff, J. (1990). Reinforced Soil Structures – Volume I: Design and Construction Guidelines (Report No.

FHWA-RD-89-043). United States. Department of Transportation. Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/1001

[4] Gowthaman, S., Nakashima, K., & Kawasaki, S. (2018). A State-of-the-Art Review on Soil Reinforcement Technology Using Natural Plant Fiber Materials: Past Findings, Present Trends and Future Directions. Materials, 11(4), 553. https://doi.org/10.3390/ma11040553

[5] Medina-Martinez, C. J., Sandoval-Herazo, L. C., Zamora-Castro, S. A., Vivar-Ocampo, R., & Reyes-Gonzalez, D. (2022). Natural Fibers: An Alternative for the Reinforcement of Expansive Soils. Sustainability, 14(15), 9275. https://doi.org/10.3390/su14159275

[6] Jui-Sheng Chou, Kuo-Hsin Yang, and Jie-Ying Lin, 2016. Peak Shear Strength of Discrete Fiber-Reinforced Soils Computed by Machine Learning and Metaensemble Methods, Publication: Journal of Computing in Civil Engineering , Volume 30, Issue 6 . https://doi.org/10.1061/(ASCE)CP.1943-5487.0000595

[7] Choudhury, D. Advances in Geotechnical Earthquake Engineering for Geo-Structures: Transformative Research into Practice. Indian Geotech J 56, 4313–4355 (2026). https://doi.org/10.1007/s40098-025-01450-w

[8] Gul, N., & Mir, B. A. (2022). Parametric study of glass fiber reinforced fine-grained soil with emphasis on microstructural analysis. International Journal of Geotechnical Engineering, 16(6), 716–728. https://doi.org/10.1080/19386362.2022.2049524

[9] Shukla, S., Sivakugan, N., & Das, B. (2009). Fundamental concepts of soil reinforcement — an overview. International Journal of Geotechnical Engineering, 3(3), 329–342. https://doi.org/10.3328/IJGE.2009.03.03.329-342

[10] Mozumder, R.S., Bhumik, M.C. & Sheikh, P. Influence of bamboo fiber reinforcement on the mechanical behavior of soft soil. Discov Civ Eng 2, 181 (2025). https://doi.org/10.1007/s44290-025-00339-0

[11] Karimah, A., Ridho, M. R., Munawar, S. S., Ismadi, Amin, Y., Damayanti, R., Lubis, M. A. R., Wulandari, A. P., Nurindah, Iswanto, A. H., Fudholi, A., Asrofi, M., Saedah, E., Sari, N. H., Pratama, B. R., Fatriasari, W., Nawawi, D. S., Rangappa, S. M., & Siengchin, S. (2021). A Comprehensive Review on Natural Fibers: Technological and Socio-Economical Aspects. Polymers, 13(24), 4280. https://doi.org/10.3390/polym13244280

[12] Aldalbahi, A., El-Naggar, M. E., El-Newehy, M. H., Rahaman, M., Hatshan, M. R., & Khattab, T. A. (2021). Effects of Technical Textiles and Synthetic Nanofibers on Environmental Pollution. Polymers, 13(1), 155. https://doi.org/10.3390/polym13010155

[13] Shalchian, M.M., Arabani, M. Application of plant-derived fibers in soil reinforcement on experimental, numerical, and case study scales: a review. Bull Eng Geol Environ 82, 19 (2023). https://doi.org/10.1007/s10064-022-03029-8

[14] Dong Zhou, Xiaoyan Tong, Hongjun Liu, Shengli Lv, T. S. Srivatsan, Xiaosheng Gao; A modified Halpin–Tsai model for predicting the elastic modulus of composite materials. AIP Advances 1 January 2024; 14 (1): 015157. https://doi.org/10.1063/5.0160256

[15] RADOSLAWL. MICHALOWSKI,1997. AN ESTIMATE OF THE INFLUENCE OF SOIL WEIGHT ON BEARING CAPACITY USING LIMIT ANALYSIS,1997 年 37 巻 4 号 p. 57-64, DOI: https://doi.org/10.3208/sandf.37.4_57

[16] Čermák, V. (1979). Heat Flow Map of Europe. In: Čermák, V., Rybach, L. (eds) Terrestrial Heat Flow in Europe. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-95357-6_1

[17] Salah Sadek, Shadi S. Najjar, and Fadi Freiha, 2010. Shear Strength of Fiber-Reinforced Sands, Publication: Journal of Geotechnical and Geoenvironmental Engineering, Volume 136, Issue 3https://doi.org/10.1061/(ASCE)GT.1943-5606.0000235

[18] Jui-Sheng Chou, Kuo-Hsin Yang , and Jie-Ying Lin, 2016. Peak Shear Strength of Discrete Fiber-Reinforced Soils Computed by Machine Learning and Metaensemble Methods, Publication: Journal of Computing in Civil Engineering. Volume 30, Issue 6 . https://doi.org/10.1061/(ASCE)CP.1943-5487.0000595

[19] Wei, G., Wang, Z., Cao, X., & Wen, J. (2025). Predictive Analysis of the Mechanical Properties of Biopolymer–Fiber-Reinforced Composite-Stabilized Soil Based on Genetic Algorithm-Optimized Back Propagation Neural Networks. Polymers, 17(16), 2176. https://doi.org/10.3390/polym17162176

[20] Onyelowe, K.C., Hanandeh, S., Kamchoom, V. et al. Mechanical properties of self compacting concrete reinforced with hybrid fibers and industrial wastes under elevated heat treatment. Sci Rep 15, 12753 (2025). https://doi.org/10.1038/s41598-025-96899-3

[21] Pham, T.A., Nadimi, S. & Sutman, M. Critical Review of Physical-Mechanical Principles in Geostructure-Soil Interface Mechanics. Geotech Geol Eng 42, 6757–6808 (2024). https://doi.org/10.1007/s10706-024-02954-7

[22] Alfio Quarteroni, Paola Gervasio, and Francesco Regazzoni, Combining physics-based and data-driven models: advancing the frontiers of research with scientific machine learning, Mathematical Models and Methods in Applied SciencesVol. 35, No. 04, pp. 905-1071 (2025), https://doi.org/10.1142/S0218202525500125

[23] Yaghoubi, E., Yaghoubi, E., Khamees, A. et al. A systematic review and meta-analysis of artificial neural network, machine learning, deep learning, and ensemble learning approaches in field of geotechnical engineering. Neural Comput & Applic 36, 12655–12699 (2024). https://doi.org/10.1007/s00521-024-09893-7

[24] Patrick Schratz, Jannes Muenchow, Eugenia Iturritxa, Jakob Richter, Alexander Brenning,2019. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data, Ecological Modelling, Volume 406, 2019, Pages 109-120, ISSN 0304-3800, https://doi.org/10.1016/j.ecolmodel.2019.06.002

[25] Dong, X., Yu, Z., Cao, W. et al. A survey on ensemble learning. Front. Comput. Sci. 14, 241–258 (2020). https://doi.org/10.1007/s11704-019-8208-z

[26] Takeru Kanazawa, Katsufumi Hashimoto, Kyoungsoo Park, Hayato Takahashi, 2025. Statistical Perspective to Time-dependent Pull-out Behavior of Discrete Fibers: A Review,Journal of Advanced Concrete Technology, 2025 年 23 巻 2 号 p. 99-115, DOI https://doi.org/10.3151/jact.23.99

[27] RADOSLAWL. MICHALOWSKI,1997. AN ESTIMATE OF THE INFLUENCE OF SOIL WEIGHT ON BEARING CAPACITY USING LIMIT ANALYSIS,1997 年 37 巻 4 号 p. 57-64, DOI: https://doi.org/10.3208/sandf.37.4_57

[28] Čermák, V. (1979). Heat Flow Map of Europe. In: Čermák, V., Rybach, L. (eds) Terrestrial Heat Flow in Europe. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-95357-6_1

[29] Zornberg JG (2002), "Discrete framework for limit equilibrium analysis of fibre-reinforced soil". Geotechnique, Vol. 52 No. 8 pp. 593–604, doi: https://doi.org/10.1680/geot.2002.52.8.593

[30] Khanduri, V., Bhardwaj, A. Prediction of Shear Strength Parameters of Fiber-reinforced Soil Using Machine Learning. Transp. Infrastruct. Geotech. 13, 185 (2026). https://doi.org/10.1007/s40515-026-00956-x

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 1, 2026, pp. 39-61
ISSN 3104-8897

[31] : Qu, J., Feng, X., Jia, X. et al. Precision zoning and dominant factor identification of soil lead using a geographical pattern interaction model and POI data. Environ Sci Eur (2026). https://doi.org/10.1186/s1230 2-026-01438-0

[32] Sen, S., Sinha, S. Tree-based machine learning models for compressive strength prediction of high-volume fly ash concrete reinforced with multiwalled carbon nanotubes. Asian J Civ Eng 26, 2915–2932 (2025). https://doi.org/10.1007/s42107-025-01351-6

[33] Hans-Georg Müller. Ulrich Stadtmüller. "Generalized functional linear models." Ann. Statist. 33 (2) 774 - 805, April 2005. https://doi.org/10.1214/009053604000001115

[34] Cyrille Prosper Ndepete, Sedat Sert, Ahmet Beycioğlu, Burak Yiğit Katanalp, Ezgi Eren, Baki Bağrıaçık, Syzmon Topolinski,2022. Exploring the effect of basalt fibers on maximum deviator stress and failure deformation of silty soils using ANN, SVM and FL supported by experimental data, Advances in Engineering Software, Volume 172, 2022, 103211, ISSN 0965-9978, https://doi.org/10.1016/j.advengsoft.2022.103211

[35] Hu, S., Mao, Z., Deng, L., Wang, Q., Liu, X., & Wang, Z. (2026). SVR-Based Framework for Predicting Stability of Circular-Failure Slopes with Small Sample Size. Mathematics, 14(12), 2074. https://doi.org/10.3390/math14122074

[36] Cagnini, H. E. L., Das Dôres, S. C. N., Freitas, A. A., & Barros, R. C. (2023). A survey of evolutionary algorithms for supervised ensemble learning. The Knowledge Engineering Review, 38, e1. doi: https://doi.org/10.1017/S0269888923000024

[37] Rew, J., Cho, Y., & Hwang, E. (2021). A Robust Prediction Model for Species Distribution Using Bagging Ensembles with Deep Neural Networks. Remote Sensing, 13(8), 1495. https://doi.org/10.3390/rs13081495

[38] Guixuan Lin, Peijian Zhou, Yanzhao Wu, Zhifeng Yao, Hongfei Wu, 2026, A heterogeneous multi-model stacking framework for predicting pump complete characteristic curves, Energy, Volume 342, 2026, 139639, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2025.139639

[39] David A. Hofmann,1997. An overview of the logic and rationale of hierarchical linear models, Journal of Management, Volume 23, Issue 6, 1997, Pages 723-744, ISSN 0149-2063, https://doi.org/10.1016/S0149-2063(97)90026-X